Are we analyzing speciation without prejudice?

Kerstin Johannesson
Department of Marine Ecology, Tjärnö, Strömstad, University of Gothenburg, Sweden

Address for correspondence: Kerstin Johannesson, Department of Marine Ecology, Tjärnö, SE-452 96 Strömstad, University of Gothenburg, Sweden. Kerstin.Johannesson@marecol.gu.se

Physical isolation has long been the null hypothesis of speciation, with exceptional evidence required to suggest speciation with gene flow. Following recent persuasive theoretical support and strong empirical examples of nonallopatric speciation, one might expect a changed view. However, a review of 73 recent empirical studies shows that when allopatric speciation is suggested, a nonallopatric alternative is rarely considered, whereas the opposite is true in studies suggesting sympatric speciation, indicating a biased treatment of different speciation models. Although increasing support for ecological speciation suggests natural selection as the most critical component of speciation, gene flow remains an issue. Methods for unbiased hypothesis testing are available, and the genetic and phylogeographic data required for appropriate tests can be generated. Focus on phylogenies and functions of individual genes have revealed strong idiosyncratic elements of speciation, such as single genes with possible allopatric origin that make significant contributions during nonallopatric phases of speciation. Hence a more complex picture of speciation is now emerging that will benefit from unbiased evaluation of both allopatric and sympatric mechanisms of speciation.

Keywords: allopatric speciation; sympatric speciation; hypothesis testing; scientific controversy

Introduction

Speciation is a fascinating part of evolution. Something in the range of 10–20 million species are current inhabitants of the globe, and these are just but a tiny fraction of all species that have been present. Hence, the total number of speciation events that have occurred is huge, most likely on the order of hundreds of millions, possibly as many as three new species per year since early Cambrian. Despite speciation being such a common phenomenon, and living and fossil evidences of its existence are all around, we have not yet agreed on its primary mechanisms. Although Darwin considered speciation to be driven by natural selection and essentially similar to the mechanism causing divergence of populations and formation of varieties and races, Dobzhansky put forward the argument that reproductive isolation forms under extrinsic barriers to gene flow. This argument was strongly supported by Mayr who suggested that extrinsic, physical barriers, and allopatric speciation was the only possible way new species could form, and this view dominated textbooks in evolution up to the early 1980s. Already at that time, however, it was clear from theoretical analyses and new empirical data that sympatric speciation could not simply be dismissed.

Today we know that pleiotropic effects, for example, via “magic traits,” as well as habitat or host-choice, relax the conditions for sympatric speciation considerably and that a variety of different contexts will enable sympatric speciation under modeled conditions. In addition, it has been acknowledged that few allopatric distributions completely prevent migration, and this raises the issue of how important rare migration events are in transferring successful mutations that will support cohesion rather than divergence of species. Moreover, physical barriers isolating incipient species presumably evolved gradually, promoting ecological divergence in parapatry across environmental gradients as an initial step. Finally, with similar types of selective pressures, genetic incompatibilities may evolve less easily than earlier believed, as directional
selection is likely to favor repeated occurrence of the same mutations in independent populations.  

Given the knowledge background and the recent addition of powerful population genetic and phylogenetic methods to test different hypotheses of speciation in any type of organism, one may ask if these methodological developments have paved the road for approaches that treat both allopatric and sympatric hypotheses of speciation as valid models to be tested. To find out if this is the case, I reviewed recent publications that proposed one or several hypotheses of speciation. For each I assessed which type of speciation (allopatric, parapatric, or sympatric) was suggested, how alternative hypotheses were treated, and what type of conceptual model, data, and tests were used to support the conclusions.

Biased treatment of allopatric and sympatric models

There has been increasing interest in biological speciation over the time period 1960–2009 (Fig. 1; early peaks may be due to stochastic variation in a small total volume of publications). Reviewing all original publications on biological speciation published very recently (Web of Science search, January 1 to November 1, 2009; N = 73) reveals a strong discrepancy in the approach researchers take to allopatric, parapatric, and sympatric models of speciation (Table 1). Only about 30% of the studies suggesting allopatric speciation (“studies suggesting allopatric speciation” subsequently = AS studies) ever discuss nonallopatric alternatives. In contrast, more than 70% of studies suggesting sympatric or parapatric speciation (“studies suggesting sympatric or parapatric speciation” subsequently = S/PS studies) consider an allopatric alternative. Focusing on studies published in high impact journals (impact factor > 4), the corresponding figures are 4 of 16 AS studies consider a nonallopatric alternative, while 9 of 11 S/PS studies discuss allopatric alternatives. Thus, there remains a strong bias in how the two groups of hypotheses are treated.

An examination of the approaches used in the different studies makes it clear that the reason nonallopatric alternatives are left out is not that the support for allopatric speciation is overwhelmingly strong: the majority of AS studies (80%) based their conclusion not on a conceptual model, but instead on circumstantial evidence, such as the current distribution of species (even when origins were quite ancient). A further complication is to show that distributional gaps between species are actually barriers against migration and gene flow (as discussed earlier).

For S/PS studies, a smaller proportion (30%) had no clear reference to a conceptual model, instead referring to unspecified observations of current distribution. An additional difference was that S/PS studies used a combination of phenotypic and genetic data three times as often as studies suggesting allopatric speciation (relying mostly on genetic data). Thus, hypotheses of speciation are quite often presented without appropriate tests or even mention of possible alternative explanations, and this tendency...


### Table 1. Publication statistics and research approaches in studies suggesting allopatric, sympatric, and parapatric speciation published from January 1, 2009 to November 1, 2009 (data from Web of Science)

<table>
<thead>
<tr>
<th>Studies suggesting allopatric speciation</th>
<th>Studies suggesting sympatric or parapatric speciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of studies</td>
<td></td>
</tr>
<tr>
<td>Journal impact &gt; 4</td>
<td></td>
</tr>
<tr>
<td>Journal impact 2–4</td>
<td></td>
</tr>
<tr>
<td>Journal impact &lt; 2</td>
<td></td>
</tr>
<tr>
<td>Alternative models considered, all journals</td>
<td></td>
</tr>
<tr>
<td>Alternative models considered, high impact journals only</td>
<td></td>
</tr>
<tr>
<td>Test of conceptual model</td>
<td></td>
</tr>
<tr>
<td>Using both genetic and ecological data</td>
<td></td>
</tr>
</tbody>
</table>

| 47 | 21 |
| 34% | 52% |
| 36% | 19% |
| 30% | 29% |
| 30% | 70% |
| 25% | 82% |
| 21% | 71% |
| 15% | 43% |

*In addition, there were five studies, dealing with groups of species, that suggested both allopatric and nonallopatric speciation. These are not included here.

A framework for how allopatric tests should be performed was outlined some years ago, but this seems not yet widely in use. In fact only three out of 47 AS studies show convincing attempts to reject alternative explanations. One study examined mechanisms of speciation in 25 species of the plant genus *Cyclamen*, and correlated the time since biogeographic events (formation of physical barriers) with the estimated time of divergence of species from phylogenetic analysis, also noting that no sister species pair currently have overlapping distributions. A comprehensive mtDNA phylogeny of 92 *Daphnia* species from seven continents looked for the coincidence of divergence of new clades and intercontinent or large-scale intracountry dispersal. They found that 42% of the speciation events took place in conjunction with colonization of new geographic areas, suggesting a minimum level of allopatric speciation. Finally, a completely different approach to test data against predictions from allopatric and nonallopatric speciation models was used by Webster who used large amounts of genome data from the human and chimpanzee genomes to compare the times of divergence of individual loci. He found strong support for allopatric speciation: the distribution of divergence times was random, indicating that genetic exchange in both coding and noncoding loci was separated by the same extrinsic barrier.

Among the S/PS studies, a large proportion (86%) have convincing (albeit not fully conclusive) data sets that often included some measure of current gene flow along with the ecological settings for different incipient species or ecotypes explaining what type of divergent selection might drive them apart. A large difference from AS studies was that in the nonallopatric case researchers dealt with ongoing speciation processes where reproductive barriers, gene flow, and ecological separation or even selection can be assessed in direct estimates, although the final result may be impossible to infer as speciation is not yet completed.

### The need for appropriate hypothesis testing

Few of the AS studies used appropriate data to test predictions from an allopatric speciation model. For example, in 19 of 47 studies, biogeographic data showing present distributions of species were used to support allopatric speciation, despite the obvious risk that current distribution may not be a reliable proxy for the distribution at the point of speciation often hundreds of thousands of years ago. In six of these AS studies, current distributions were overlapping and hypothetical barriers were suggested to have been in existence during earlier periods. Finally, in 16 of the AS studies no suggestions were made for the mechanisms that initiated divergence, but allopatric speciation was suggested on the basis of the current genetic separation of species.

is much stronger in AS studies than in S/PS studies. Somewhat surprisingly, the same tendency is strong even in high-impact journals.

### Intrinsic problems in comparisons of modes of speciation

As allopatric speciation is usually inferred from events distant in the past, and presumably under conditions that cannot be deduced from observing the resulting species in their current context, we may have to accept a poor resolution of the mechanisms

---

that initiated separation into later independent evolutionary lineages (The Spyglass perspective\textsuperscript{36}). In contrast, sympatric speciation must be investigated at an incipient stage (The Magnifying Glass perspective\textsuperscript{29}), or soon thereafter, as the accumulation of genetic differences after the completion of sympatric speciation will efficiently hide the footprints of the mechanisms involved in isolation (with the exception of speciation through polyploidy). Although we cannot be sure that incipient speciation will eventually be completed, we will have to accept that this “point of no return” may be hard to define; the best we can do is probably to examine cases where reproductive isolation owing to ecological forces is almost as strong as between distinct species.\textsuperscript{29,30} Given that in most cases very different approaches are applied to study allopatric and sympatric speciation, quantitative comparisons may be too complex, with the exception of genomic approaches like the one used by Webster,\textsuperscript{28} that are today made possible by the introduction of next-generation DNA sequencing.

What is generally agreed is that it is hard to conceive either a strict sympatric setting (completely random mating) or a strict allopatric setting (no gene flow) in nature.\textsuperscript{21,31,32} Moreover, most allopatric scenarios may very well be preceded by a preallopatric stage of parapaty, as physical barriers to gene flow such as the raising of mountain ridges, glacial formation, and changing oceanic currents is unlikely to be instantaneous. Perhaps then, we should consider parapatric speciation as the most likely type of speciation in nature.\textsuperscript{21} Such a conclusion is, however, rejected in the recent literature; only two of 73 reviewed studies suggested parapatric speciation. An additional complicating factor is the disagreement on precise definitions of allopatry, parapaty, and sympatry,\textsuperscript{31} and the issue of whether these concepts only describe geographic settings, or if they also indicate levels of gene flow.\textsuperscript{33}

It is hence not a trivial problem to separate allopatric and sympatric speciation mechanisms in nature, and perhaps it would be better to not stress such differences at all. One simple way to avoid the problem is to leave out the spatial setting of speciation and instead focus on the forces that create genetic differences.\textsuperscript{13,34} Schluter\textsuperscript{35} advocates differentiating between speciation driven by natural selection and mutation-order driven speciation, the latter caused by sexual conflicts, cytoplasmic male sterility, meiotic drive, and similar. He thereby omits allopatric situations from the discussion, and hence also the role of genetic drift. Thus, even if it is interesting to compare deterministic (selection) and stochastic (mutations) forces driving speciation in a parapatric or sympatric setting, such an approach cannot cast light on the process of speciation in the absence of gene flow.

**A proper representation of Tree of Life**

Today particular groups of organisms, for example, freshwater fishes and host race forming species, are overrepresented among sympatric speciation cases studied, while examples of allopatric speciation involving island populations have a bias toward bird and reptile species.\textsuperscript{36,37} New and powerful genome screening technologies can generate comprehensive genetic data for any type of organism and help to provide new data to increase information about speciation mechanisms over parts of the Tree of Life that are currently poorly represented. Another option will be to analyze ancient genomes by extracting DNA from fossils,\textsuperscript{38,39} and this will in the future possibly give us a tool to estimate gene flow and differentiation back in time between lineages that we now know are distinct species.

One already successful genome approach is the coalescence-based modeling of past isolation and migration (gene flow) patterns (the IM model\textsuperscript{40}). Under this approach, current phylogenetic separation is compared with models of divergence with or without gene flow (sympatric/parapatric vs. allopatric models). It is also possible to estimate the pattern of gene flow over time, allowing separation of allopatric divergence followed by secondary contact, and introgression from divergence during ongoing periods of gene flow.\textsuperscript{16} Nice examples include the use of the IM model to support parapatric speciation of cave salamanders from spring salamanders,\textsuperscript{41} and incipient allopatric speciation in a Chinese perennial herb.\textsuperscript{42} Although, there are some concerns with how well the assumptions of the IM models are usually met,\textsuperscript{43} further development of coalescence-based modeling is likely to be an important contribution to speciation mechanism hypothesis testing.

A related approach, based on similar coalescence assumptions, is to compare large parts of genome sequences under the assumption that selected and neutral genes should show equal distributions of
divergence times if there had been no gene flow since the first separation of the two lineages. Analyzing nearly a million nuclear sites of the human and chimpanzee genomes, Webster found support for complete allopatric divergence of these two primate lineages.

**Mixed-mode speciation**

What becomes more and more obvious is that a particular speciation event may involve both sympatric, parapatric, and allopatric phases. The role of reinforcement in speciation is one example when an initial allopatric separation generates genetic differences that lead to partial reproductive isolation; following secondary contact of the incipient species, selection against hybrids reinforce the barrier to completion. Interestingly, it is also possible to conceive an alternative order of events: Via divides speciation into a first phase during which divergent selection separates two incipient and nonallopatric species under the development of prezygotic barriers. As a consequence of these barriers, speciation proceeds under what Via calls “ecological allopatry” during which genetic incompatibility evolves as a by-product of the impeded gene flow.

Speciation resulting from allopatric divergence followed by secondary introgression and gene flow may be hard to distinguish from sympatric and parapatric divergence both in principle and in nature, because secondary gene flow obscures earlier divergence during allopatry and may result in a phylogenetic pattern very similar to what is expected under divergence with gene flow. Although idiosyncratic (site-specific) signals of divergence will support local origin of differences, it is obvious that in many cases separation may involve components of both allopatric and sympatric divergence. It has, for example, been shown that populations that diverge in the presence of gene flow may do so because of ancestral genes with a possible allopatric origin. Such genes are likely to be present in low frequencies in the ancestral species and, upon encountering a new niche, promote adaptation and formation of new ecotypes that may even be reproductively isolated from the ancestral species. An interesting example is the chromosome inversion found in members of *Rhagoletis* flies that originated in Mexico but became important to ecological divergence through host shifts in North America at a much later stage. A similar striking example is the *Eda* gene that is of ancient origin and present in recessive form in heterozygotes in ~1% of marine sticklebacks. This gene has played a major role when freshwater ecotypes of sticklebacks have evolved repeatedly in postglacial streams and lakes. Freshwater populations have fewer skeletal plates, and fishes of these populations are fixed for the recessive *Eda* allele, indicating that this gene is under strong directional selection. Hence in both of these situations, ancestral genes of allopatric origin are part of the standing genetic variation that promotes speciation with gene flow.

A somewhat contrasting mechanism that has recently been highlighted by new empirical results is the hitchhiking of neutral genes close to a gene under selection. When relatively large parts of the genome (around 5%) are affected by divergent selection, as found in ecotypes of marine snails and lake whitefish, and in host races of pea aphids, divergent selection may interfere with patterns of variation in neutral loci that would otherwise reflect past levels of gene flow. The issue currently discussed is how large the affected part of the genome surrounding a selected loci is. This begs for more detailed studies that investigate separately the signals of different parts of the genome, such as adaptive genes, their linked noncoding regions, and unlinked noncoding regions. In sum, mixed-mode speciation, as suggested by earlier authors, emerges as a quite likely scenario of speciation and, along with parapatric speciation, is likely much more common than is currently reflected in the literature.

Nevertheless, a crucial step is to treat different models of speciation with similarly strict approaches of hypothesis testing. Therefore, the dearth of discussion of alternative explanations in the majority of AS studies does not augment support for this model in the long run. Hopefully, new approaches such as combining coalescence modeling and genome data, or comparing phylogenetic information from coding and noncoding parts of the genomes, are promising not only to contribute new data but also as they treat allopatric and nonallopatric alternatives without prejudice. Quite obviously, the increased power of detecting potential mechanisms of speciation (separately assessing “speciation genes,” as one example) will increase the complexity of the picture that emerges but nevertheless will provide us with a more thorough insight into “the mystery of the mysteries,” and it seems quite likely that in the
end, the allopatry–sympathy controversy may conclude that most speciation events have elements of both.

Acknowledgments

Inspirng discussions with among others, Olof Leimar, Roger Butlin, Carl André, Marina Panova, Petri Kemppainen, Emilio Rolán-Alvarez, Humberto Quesada and Jerry Coyne provide a background to this work. I thank Carl Schlichting for the invitation to write this review/discussion paper and Jeff Feder for numerous insightful comments that helped a lot to improve it. This work was performed within the University of Gothenburg Linnaeus Centre for Marine Evolutionary Biology (http://www.cemeb.science.gu.se).

Conflicts of interest

The author declares no conflict of interest.

References